

Introduction to Ontologies in Health and Biology

(ICF Ontology and Protégé workshop 2008)

Nigam Shah

Stanford University
nigam@stanford.edu



Data explosion in the life sciences



- Sequence information
 - The first data type to be available in large amounts
 - Has had the maximum time to be standardized
 - FASTA format is the most popular
- Expression information
 - Recent rise in abundance
- Transcription factor binding information
 - High throughput available in yeast
- Protein-Protein interaction information
 - Relatively recent rise in availability.
 - ChIP, array based.
- Past knowledge, traditional experiments, published papers.

So many biological databases, so little time



- More than 1000 different databases!
- Some biological databases:

AATDB, AceDb, ACUTS, ADB, AFDB, AGIS, AMSdb, ARR, AsDb, BBDB, BCGD, Beanref, Biolmage, BioMagResBank, BIOMDB, BLOCKS, BovGBASE, BOVMAP, BSORF, BTKbase, CANSITE, CarbBank, CARBHYD, CATH, CAZY, CCDC, CD4OLbase, CGAP, ChickGBASE, Colibri, COPE, CottonDB, CSNDB, CUTG, CyanoBase, dbCFC, dbEST, dbSTS, DDBJ, DGP, DictyDb, Picty_cDB, DIP, DOGS, DOMO, DPD, DPlnteract, ECDC, ECGC, EC02DBASE, EcoCyc, EcoGene, EMBL, EMD db, ENZYME, EPD, EpoDB, ESTHER, FlyBase, FlyView, GCRDB, GDB, GENATLAS, Genbank, GeneCards, Genline, GenLink, GENOTK, GenProtEC, GIFTS, GPCRDB, GRAP, GRBase, gRNAsdb, GRR, GSDB, HAEMB, HAMSTERS, HEART-2DPAGE, HEXAdb, HGMD, HIBD, HIDC, HIVdb, HotMolecBase, HOVERGEN, HPDB, HSC-2DPAGE, ICN, ICTVDB, IL2RGbase, IMGT, Kabat, KDNA, KEGG, Klotho, LGIC, MAD, MaizeDb, MDB, Medline, Mendel, MEROPS, MGDB, MGI, MHCPEP5, Micado, MitoDat, MITOMAP, MJDB, MmtDB, Mol-R-U, MPDB, MRR, MutBase, MycDB, NDB, NRSUB, O-lycBase, OMIA, OMIM, OPD, ORDB, OWL, PAHdb, PatBase, PDB, PDD, Pfam, PhosphoBase, PigBASE, PIR, PKR, PMD, PPDB, PRESAGE, PRINTS, ProDom, Prolysis, PROSITE, PROTOMAP, RatMAP, RDP, REBASE, RGP, SBASE, SCOP, SeqAnaiRef, SGD, SGP, SheepMap, Soybase, SPAD, SRNA db, SRPDB, STACK, StyGene, Sub2D, SubtiList, SWISS-2DPAGE, SWISS-3DIMAGE, SWISS-MODEL Repository, SWISS-PROT, TelDB, TGN, ImRDB, TOPS, TRANSFAC, TRR, UniGene, URNADB, V BASE, VDRR, VectorDB, WDCM, WIT, WormPep, YEPD, YPD, YPM, etc !!!!

3

More data is good, what's the problem?



- Too unstructured:
 - from a variety of incompatible sources
 - no standard naming convention
 - each with a custom browsing and querying mechanism
 - and poor interaction with other data sources
- Difficult to use and understand the available data, information and knowledge

4

Ontologies to the rescue



- Ontologies provide **formal specification** of how to represent objects, concepts and relationships among them
- Ontologies provide a **shared understanding [language]** for communicating biological information
- Ontologies **overcome the semantic heterogeneity** commonly encountered in biomedical databases
- Ontologies are interpretable by humans and by computer programs.

5




THE NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY

[Home](#) | [Contact](#)

[About](#) [Resources](#) [News and Events](#) [Education](#) [Dissemination](#) [Collaboration](#)


*Advancing biology and medicine with tools and methodologies
for the structured organization of knowledge.*

[Resources](#)

[Education](#)


[Dissemination](#)

[Driving Biological Projects](#)

[Collaboration](#)

☒ Search bioontology.org ☐ Search the web

Quick Links

-  [Center Tools](#)
-  [BioPortal](#)
-  [Working With The Center](#)

Current News

May 1, 2008 - New Article Published: [Towards a richer description of our complete collection of genomes and metagenomes: the Minimum Information about a Genome Sequence \(MIGS\) specification](#), Field, et al. *Nature Biotechnology*, 26(5): 541 - 547, May 2008.


February 29, 2008 - The The NCBO [BioPortal Prototype](#) is now available to the community for testing and feedback!

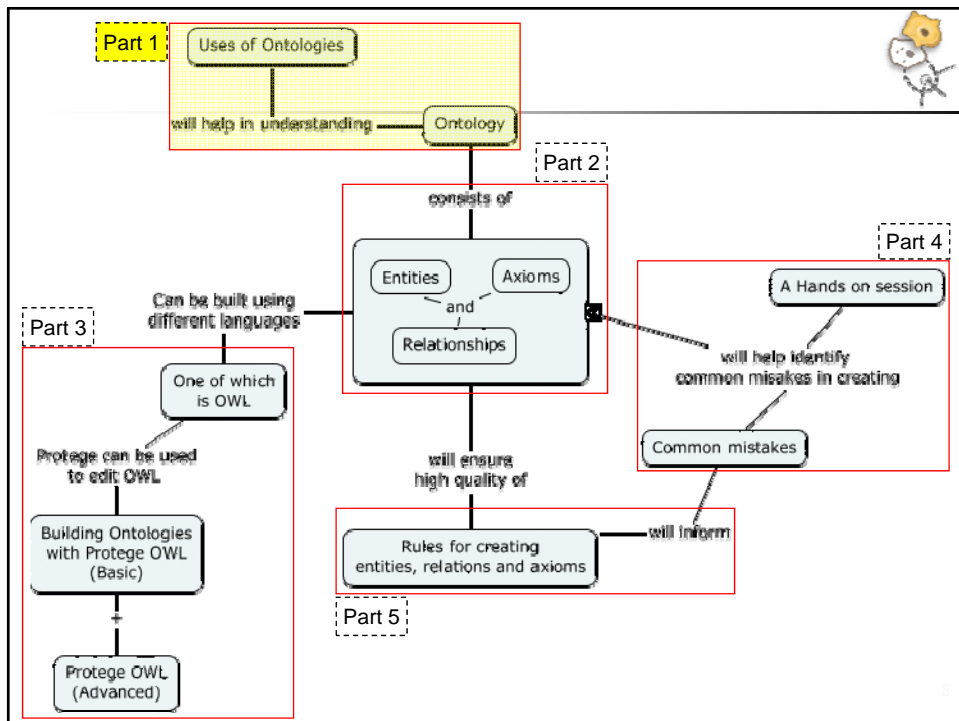
January 1, 2008 - New Article Published: [Biomedical ontologies: a functional perspective](#), D. L. Rubin, N. H. Shah, N. F. Noy. *Briefings in Bioinformatics* January 2008, 9(1):75-90.

December 8, 2007 - New Article Accepted, In Press. [Translating the Foundational Model of Anatomy into OWL](#), N. F. Noy, D. L. Rubin. *Journal of Web Semantics* In Press, Corrected Proof. Available online 8 December 2007.

Copyright Stanford University 2006

6

 NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY <i>BioPortal</i>						
Browse Search Align						
Ontologies						
View By: Category						
Select Item and ... Download Visualize Submit Ontology						
Expand All Collapse All						
Ontologies						
Select Focus	Ontology	Knowledge Representation Language	Foundry	Current Version	Release Date	Version Status
<input type="radio"/>	Ontologies					
<input type="radio"/>	Anatomy					
<input type="radio"/>	BRENDA tissue / enzyme source	OBO Text	No	1.1	5/3/2006	Production
<input type="radio"/>	Cell type	OBO Text	Yes	1.1	5/3/2006	Pre-Production
<input type="radio"/>	Drosophila gross anatomy	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>	Mosquito gross anatomy	OBO Text	No	1.1	5/3/2006	Production
<input type="radio"/>	Gross Anatomy					
<input type="radio"/>	Animal Gross Anatomy					
<input type="radio"/>	Fish Anatomy					
<input type="radio"/>	Medaka fish anatomy and development	OBO Text	Yes	1.1	5/3/2006	Pre-Production
<input type="radio"/>	Zebrafish anatomy and development	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>	Human Developmental Anatomy					
<input type="radio"/>	Human developmental anatomy, abstract version	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>	Human developmental anatomy, timed version	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>	Mouse Anatomy					
<input type="radio"/>	Mouse gross anatomy and development	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>	Mouse adult gross anatomy	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>	Microbial Anatomy					
<input type="radio"/>	Dictyostelium discoideum anatomy	OBO Text	Yes	1.1	5/3/2006	Pre-Production
<input type="radio"/>	Fungal gross anatomy	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>	Plant Anatomy					
<input type="radio"/>	Maize gross anatomy	OBO Text	Yes	1.1	5/3/2006	Pre-Production
<input type="radio"/>	Cereal plant gross anatomy	OBO Text	Yes	1.1	5/3/2006	Retired
<input type="radio"/>	Chemical					
<input type="radio"/>	Chemical entities of biological interest	OBO Text	Yes	1.1	5/3/2006	Retired
<input type="radio"/>	Physico-chemical methods and properties	OBO Text	No	1.1	5/3/2006	Pre-Production
<input type="radio"/>	Physico-chemical process	OBO Text	No	1.1	5/3/2006	Pre-Production



Uses of ontologies



1. Naming “things”

- Reference ontologies
- Controlled terms for annotating “things”

2. As a data exchange format
3. Define a knowledgebase schema
4. Computer reasoning over data
5. Driving NLP
6. Information integration

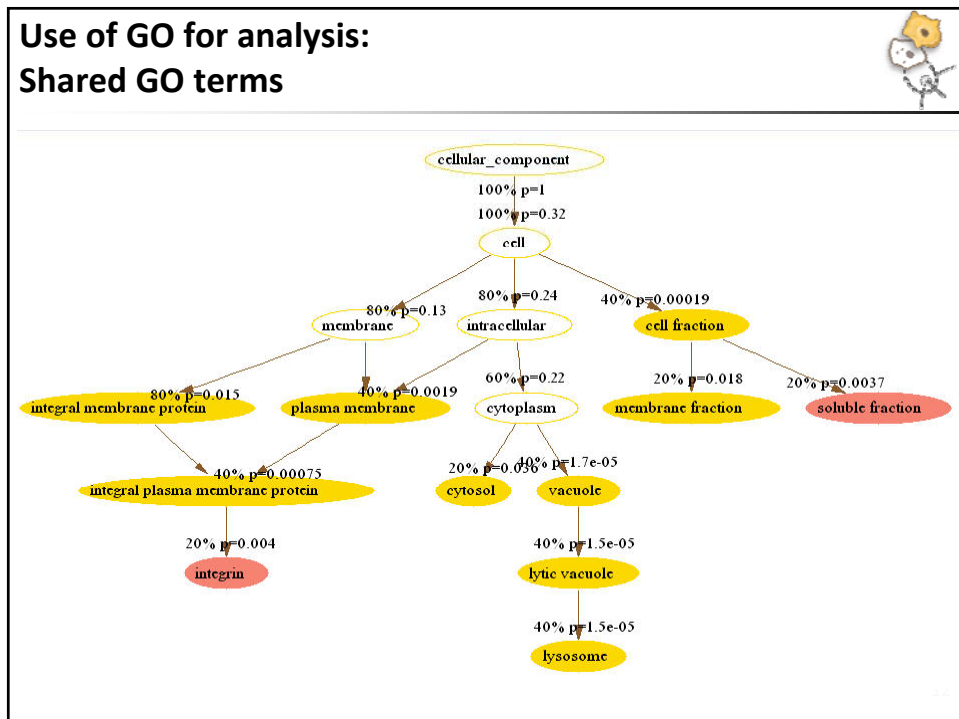
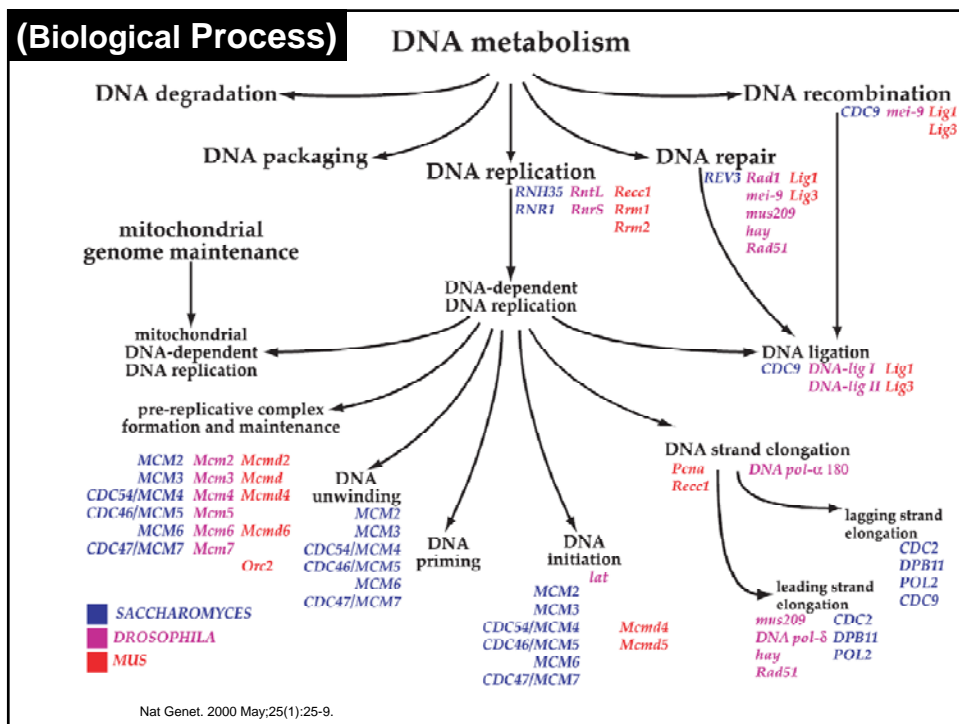
9

The Gene Ontology www.geneontology.org



- The Gene Ontology (GO) project is an effort to provide **consistent descriptions of gene products.**
- The project began as a collaboration between **three model organism databases:**
 - FlyBase (*Drosophila*)
 - *Saccharomyces* Genome Database (SGD)
 - Mouse Genome Database (MGD)
- GO creates terms for:
 - Biological Process
 - Molecular Function
 - Cellular Component

10



MESH = Medical Entity Subject Headings

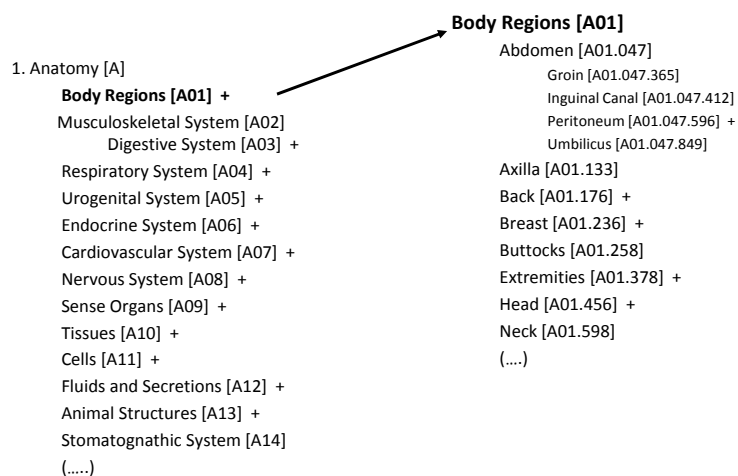


- Controlled vocabulary for indexing biomedical articles
- 19,000 “main headings” organized hierarchically
- Implicit semantics of parent-child relationships
- Multiple inheritance
- List of subheadings attached to main headings as modifiers

Copyright Stanford University 2006

13

MeSH Subtrees



14

MeSH Headings in an article



MH - Adult MH - Antipsychotic
 Agents/**pharmacology**/***therapeutic use**
 MH - **Comparative Study** ← **Supplementary heading**
 MH - Dose-Response Relationship, Drug
 MH - **Female**
 MH - Genotype
 MH - **Human** ← **Minor heading**
 MH - **Male** ← **Major heading**
 MH - Pharmacogenetics
 MH - Polymorphism (Genetics)/***genetics** ← **Qualifier**
 MH - Prognosis
 MH - Psychiatric Status Rating Scales
 MH - Receptors, Serotonin/**drug effects**/***genetics**
 MH - Risperidone/**pharmacology**/***therapeutic use**
 MH - Schizophrenia/**diagnosis**/***drug therapy**/**genetics**
 MH - Schizophrenic Psychology
 MH - **Support, Non-U.S. Gov't**
 MH - Treatment Outcome

45

Use of MeSH for Information Retrieval



NCBI PubMed National Library of Medicine NLM

Entrez PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed for Computational Biology [MH] AND Medical Informatics [MH] Go Clear

Limits Preview/Index History Clipboard Details

Display Summary Show: 20 Sort Send to Text

"Computational Biology [MH] AND Medical Informatics [MH]" of 183 Next

1: Zhang DL, Li YD, Ji L. [Correction of five different types of errors of model REFSEQs appeared in NCBI human gene database only by using two novel human genes C17orf32 and ZNF362]. Yi Chuan Xue Bao. 2004 Apr;31(4):325-34. Chinese. PMID: 15487498 [PubMed - indexed for MEDLINE] Related Articles, Links

2: Chen Y, Kortemme T, Robertson T, Baker D, Varani G. A new hydrogen-bonding potential for the design of protein-RNA interactions predicts specific contacts and discriminates decoys. Nucleic Acids Res. 2004 Sep 30;32(17):5147-62. Print 2004. PMID: 15459285 [PubMed - indexed for MEDLINE] Related Articles, Links

3: Gordon PM, Senses CW. Osprey: a comprehensive tool employing novel methods for the design of oligonucleotides for DNA sequencing and microarrays. Nucleic Acids Res. 2004 Sep 29;32(17):e133. PMID: 15456895 [PubMed - indexed for MEDLINE] Related Articles, Links

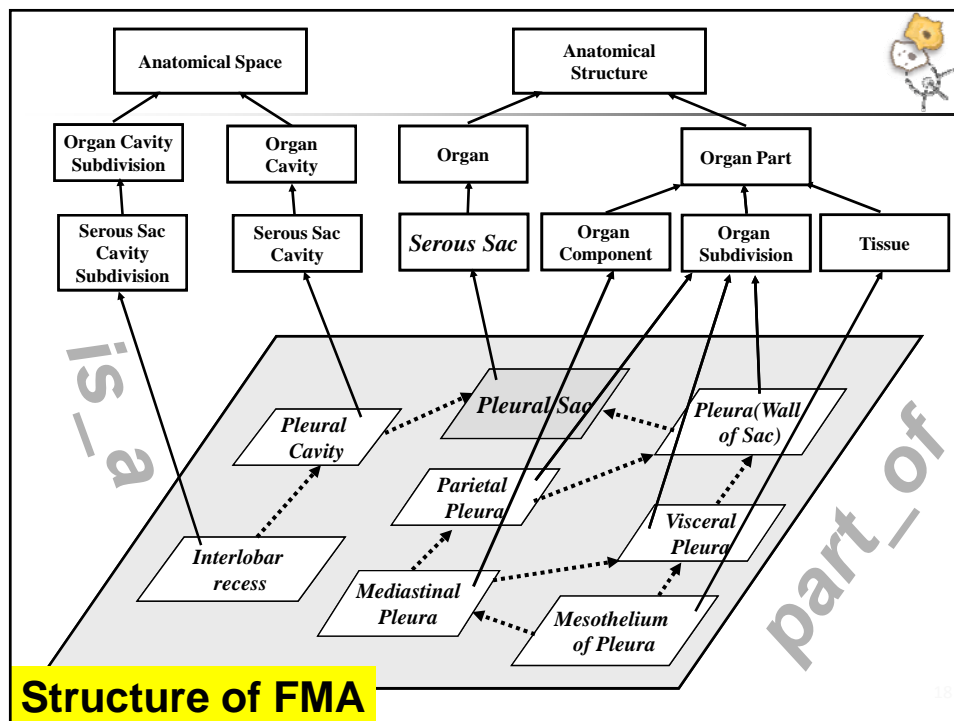
4: Wood AP, Aunikko JP, Kelly DP. A challenge for 21st century molecular biology and biochemistry: what are the causes of obligate autotrophy and methanotrophy? FEMS Microbiol Rev. 2004 Jun;28(3):335-52. Review. PMID: 15449607 [PubMed - indexed for MEDLINE] Related Articles, Links

Entrez PubMed Overview Help | FAQ Tutorial New/Noteworthy E-Utilities PubMed Services Journals Database MeSH Database Simple Citation Matcher Batch Citation Matcher Clinical Queries LinkOut Cubby Related Resources Order Documents NLM Catalog



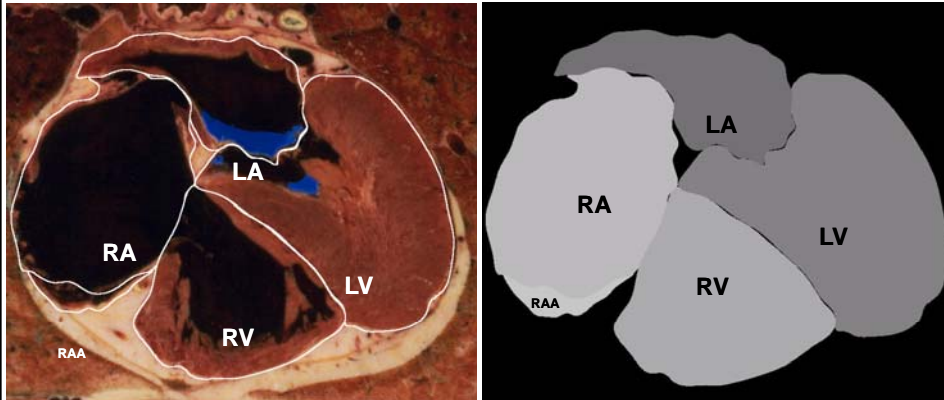
- Long-term project at University of Washington to create a comprehensive ontology of human anatomy
- 72K concepts, 1.9M relationships
- Rich semantics

37



38

Use of FMA: Image annotation



- Images possess no knowledge of their contents
- FMA-based image annotation provides that knowledge

19

Uses of ontologies



1. Naming “things”
- 2. As a data exchange format**
3. Define a knowledgebase schema
4. Computer reasoning over data
5. Driving NLP
6. Information integration

20

MGED Ontology www.mged.org



- Provides **standard terms** for annotation of microarray experiments
 - Enables **unambiguous descriptions** of how the experiment was performed
 - Enables **structured queries** of elements of the experiments

21

MGED Ontology Browser



MGED_T Taxonomy

- ExternalConcepts
- MGEDOntology
 - DataType
 - DeprecatedTerms
 - MGEDCoreOntology
 - ArrayDesignPackage
 - ArrayPackage
 - AuditAndSecurityPackage
 - BioAssayPackage
 - BioMaterialPackage
 - Action
 - BioMaterial
 - BioMaterialCharacteristics**
 - Age
 - BioSourceProvider
 - BioSourceType
 - Biometrics
 - CellLine
 - CellType
 - ChromosomalAberrationClassification
 - ClinicalTest
 - ClinicalTestType
 - ClinicalTreatment
 - DevelopmentalStage
 - DiseaseStaging
 - DiseaseState
 - EnvironmentalHistory
 - GeneticModification
 - Histology
 - Individual
 - IndividualGeneticCharacteristics
 - Organism

Concept Details

Age [Generate URI](#)

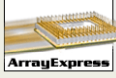
Identifiers:	
name	Age
code	C556
Roles:	
has_ID	string
has_initial_time_point	InitialTimePoint
has_maximum_measurement	Measurement
has_measurement	Measurement
is_user_defined	BibliographicReference
is_user_defined	Contact
Properties:	
Concept_Type	mged_class
Preferred_Name	Age
DEFINITION	The time period elapsed since an identifiable point in the life cycle of an organism. If a developmental stage is specified, the identifiable point would be the beginning of that stage. Otherwise the identifiable point must be specified such as planting.
Superc Concepts	
BioMaterialCharacteristics	

http://nciterns.nci.nih.gov/priv_mged_o/Connect.do

22

USE OF MGED ONTOLOGY: ArrayExpress Query form





You are logged in as guest! [Login »](#) **ArrayExpress** (392 Experiments with 11676 Hybs, 294 Arrays) [Help](#)

User **guest**, your query for Experiments
with species = **Homo sapiens**
with experiment type = **co expression**

produced
0
matches

Query for Experiments

Give an experiment **accession number** for example E-MANP-2, [Query »](#)

or fill out some of the following fields to get a list of matching experiments:

Species Homo sapiens <input type="button" value="v"/>	Author <input type="text"/>	Laboratory <input type="text"/>
Array accession number <input type="text"/>	Array design name <input type="text"/>	Array provider <input type="text"/>
Experiment type < any type > <input type="button" value="v"/>	Experimental Factors < any factor > <input type="button" value="v"/>	Description contains the word <input type="text"/>

23

Uses of ontologies



1. Naming “things”
2. As a data exchange format
- 3. Define a knowledgebase schema**
4. Computer reasoning over data
5. Driving NLP
6. Information integration

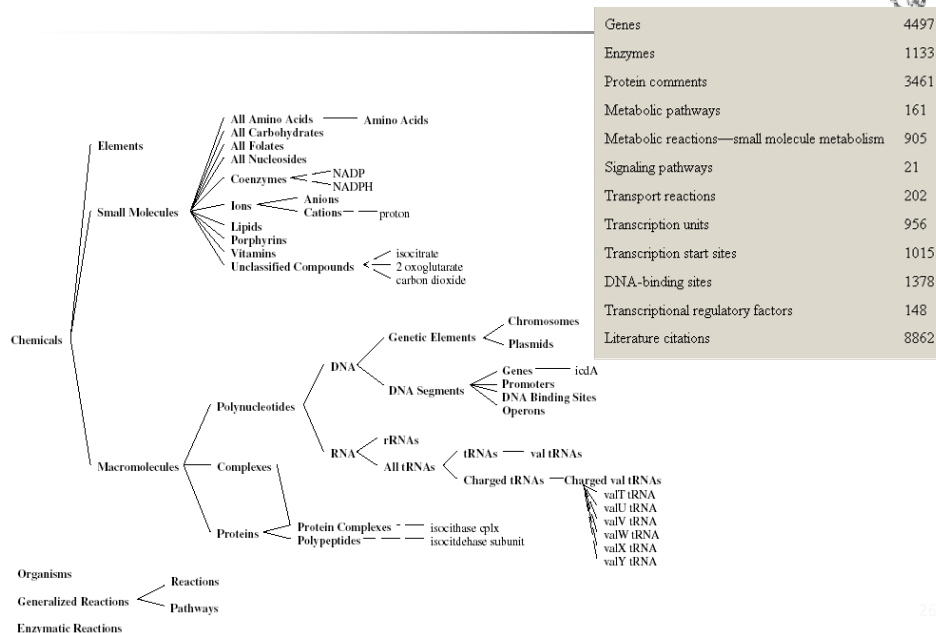
24



- The EcoCyc database is a comprehensive source of information on Escherichia coli K12.
- The mission for EcoCyc is to **contain both computable descriptions** of, and **detailed comments** describing, all genes, proteins, pathways and molecular interactions in E.coli.
- Through ongoing manual curation, extensive information has been extracted from 8862 publications and added to Version 8.5 of the EcoCyc database

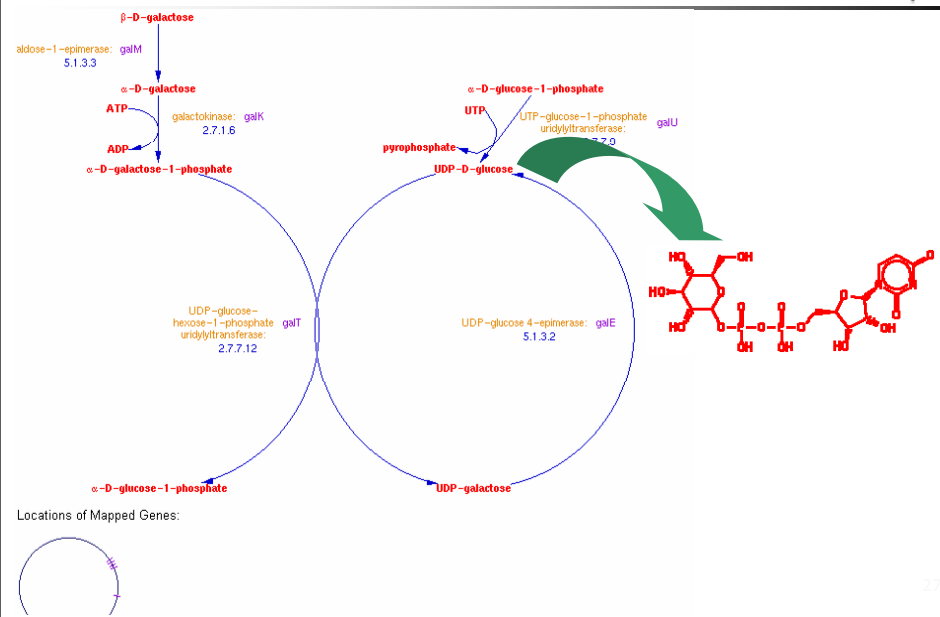
25

The EcoCyc ontology



26

Using the EcoCyc Knowledgebase



27

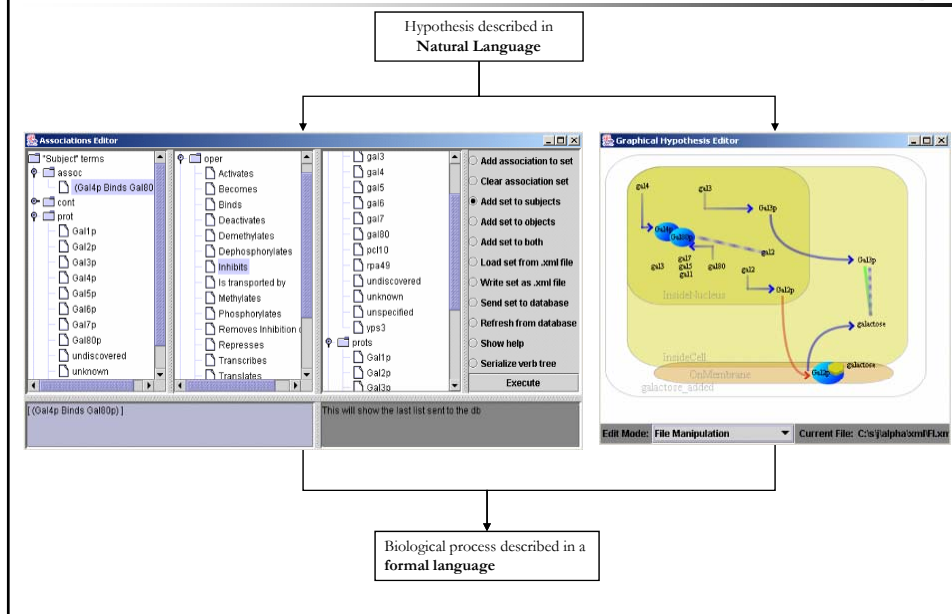
Uses of ontologies



1. Naming “things”
2. As a data exchange format
3. Define a knowledgebase schema
- 4. Computer reasoning over data**
5. Driving NLP
6. Information integration

28

HyBrow - interfaces



Evaluating an hypothesis



Query Hybrow [\[Submit a Sample file\]](#)
connected to localhost at Mon Jul 26 13:54:06 2004

First agent --- negation status --- Operator --- Second agent

gal1 --- Acetylate --- gal1

In Wt Strain This event occurs on/in mem

The Event is associated with:

presence of
absence of
only if
other

Submit Single Event

OR

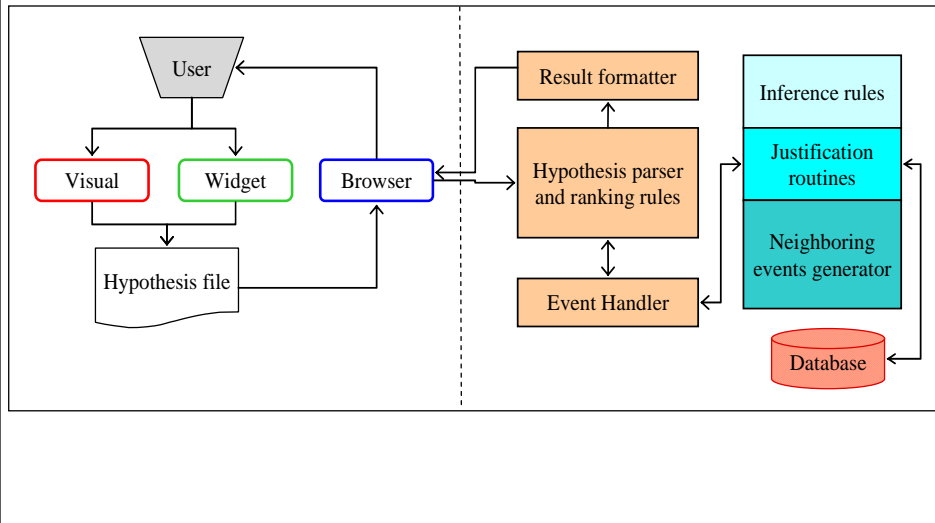
Upload a Hypothesis file for Hybrow: [\[What is a hypothesis file?\]](#)

Browse...

Submit File [\[See a Sample file\]](#) [\[Submit a Sample file\]](#)

ev0 = Gal2p transports galactose in mem in wt
ev1 = galactose activate Gal3p in wt in cyt
ev2 = Gal3p binds_to_promoter gal1 in wt in nuc
ev3 = Gal3p induce gal1 in presence_of galatose in wt in nuc
hy1 = (ev0+ev1) and (ev2+ev3)

Evaluating an hypothesis



Screen shot of the output



Event Validation

ev0 = Gal2p transports galactose in mem in wt
Table: 1Gal2p transports galactose Wt mem ([R9255129](#))
Annot: 1Gal2p: galactose transporter, .

ev1 = galactose activate Gal3p in cyt in wt
Table: 1galactose activate Gal3p Wt cyt ([Q5324798](#))

ev2 = Gal3p Binds_to_promoter gal1 in nuc in wt
Outo: 1Agent b has to be gene for Binds_to_promoter
Annot: 1Gal3p: transcriptional activator, , nucleus

ev3 = Gal3p induce gal1 in nuc in wt in presence_of galatose
Outo: 1 Agent b has to be gene for induce
Table: 1 Gal3p induce gal1 Wt nuc ([20266277](#))
Annot: 1Gal3p: transcriptional repressor, , cytoplasm

h₁ = (ev0+ev1) and (ev2+ev3)
ev0 = Gal2p transports galactose in mem in wt
ev1 = galactose activate Gal3p in cyt in wt
ev2 = Gal3p Binds_to_promoter gal1 in nuc in wt
ev3 = Gal3p induce gal1 in nuc in wt in presence_of galatose
Total supp:9, Total contr:1

	8	9	10	11	12
C					
O					
N					
I					
R					

SUPPORT

Changed ev2 to: Gal4p Binds_to_promoter gal1 in nuc in wt
Changed ev3 to: Gal4p induce gal1 in nuc in wt in presence_of galatose

Explanation

Holding the mouse on a neighbouring hypothesis (b1) shows what event was replaced to create it

Uses of ontologies



1. Naming “things”
2. As a data exchange format
3. Define a knowledgebase schema
4. Computer reasoning over data
- 5. Driving NLP**
6. Information integration

33

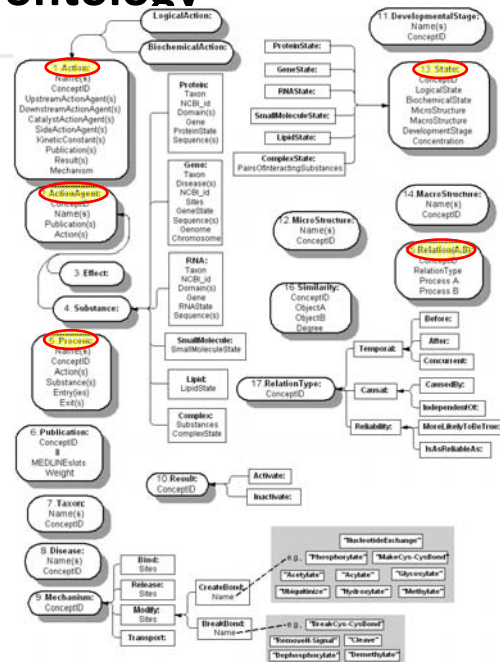
Geneways: geneways.cu-genome.org



- **Common tasks for NLP**
 - automated **selection of articles** pertinent to molecular biology,
 - automated **extraction of information** using natural-language processing,
 - **generation of specialized knowledge bases** for molecular biology.
- GeneWays is an integrated system that combines several such subtasks.
 - It **analyzes interactions** between molecular substances, **drawing on multiple sources of information** to infer a consensus view of molecular networks.

34

Geneways ontology



35

Use of Geneways ontology

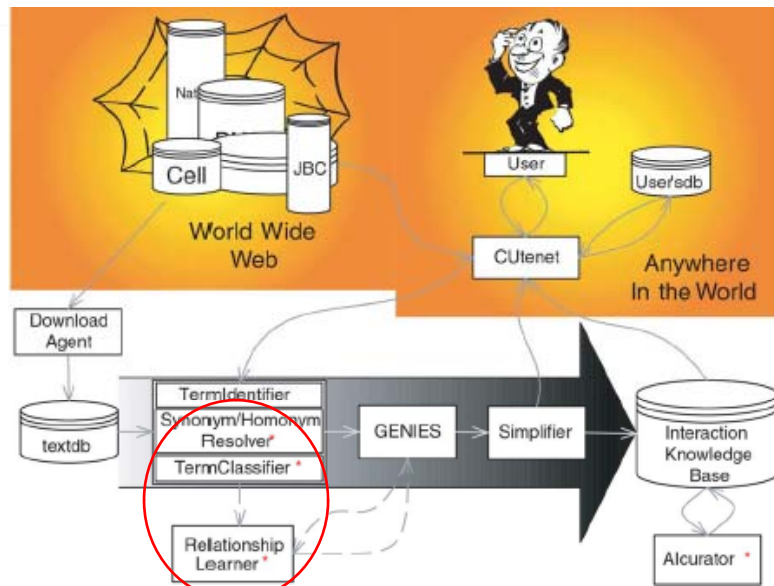


Fig. 1. A simplified view of GeneWays system.

36

Uses of ontologies



1. Naming “things”
2. As a data exchange format
3. Define a knowledgebase schema
4. Computer reasoning over data
5. Driving NLP
- 6. Information integration**

37

TAMBIS



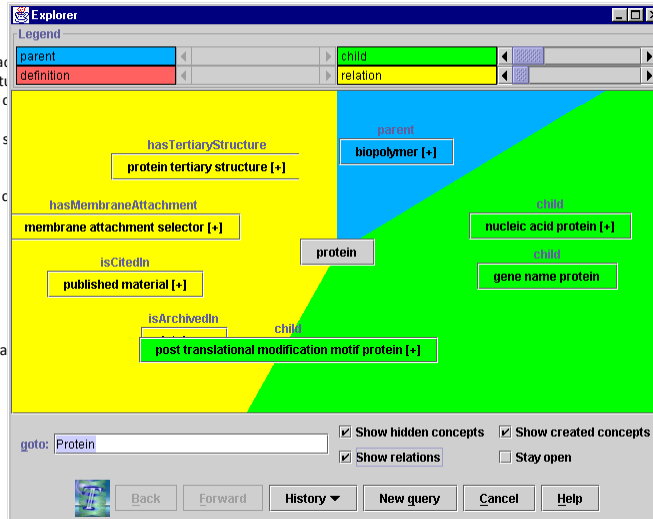
- **Transparent Access to Multiple Bioinformatics Information Sources**
- Motivation: Difficult to query distributed bioinformatics resources
- Concept:
 - Use an ontology to manage presentation and usage of diverse resources
 - Provide homogenizing layer over numerous heterogeneous databases & tools
 - Provide common, consistent query interface

38

TAMBIS browser

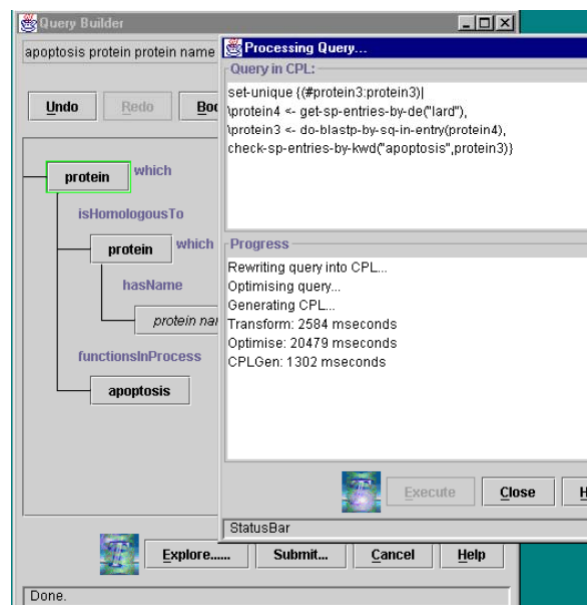


Is archived in: database
 is cited in: published material
 has membrane attachment: membrane attachment
 has tertiary structure: protein tertiary structure
 has cellular location: organelle, membrane, cytoplasm
 has name: gene name, protein name
 has secondary structure: protein secondary structure
 has identifier: identifier
 has accession number: accession number
 functions in process: biomolecular process, catalytic activity
 is bound by: protein, binding site
 binds: protein
 is homologous to: protein, nucleic acid
 is coded for by: exon, mRNA, DNA
 is translated from: DNA, mRNA
 catalyzes: reaction
 has organism classification: species
 has modification: post translational modification
 forms part of: protein complex
 has prosthetic group: prosthetic group
 is expressed in organ: organ
 has component: chemical binding site, post translational modification motif, domain
 is component of: protein complex
 is encoded by: gene
 has sequence: sequence

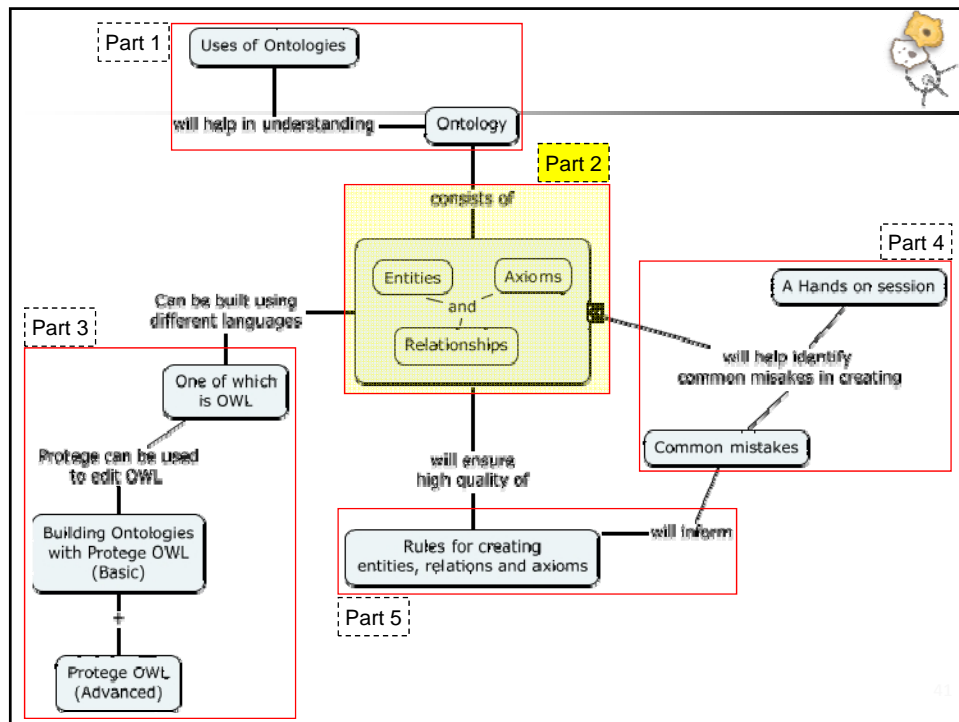


39

Query Result

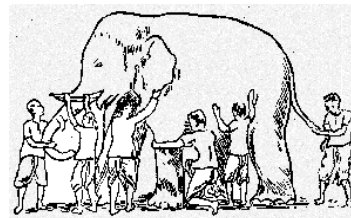


40



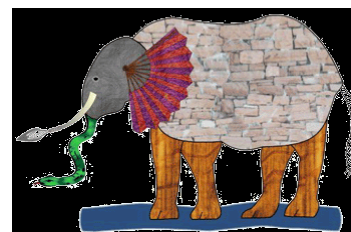
Various meanings of Ontology

Philosophy: Ontology is the study of what **entities** and what **types** of entities exist in **reality** and the **relationships** that exist between them.



AI: An ontology is an explicit specification of **concepts** & relationships that can exist in a **domain of discourse**.

IT: an ontology is a **data model** that represents a domain and is used to reason about the **objects in that domain** and the relations between them.



The common ground...



Ontology = A specification of entities (or concepts), relations, instances and axioms in an area of study.

43



ENTITIES

Representing entities



- | | |
|--|---|
| 1. Physical Reality | A. The reality on the side of the patient |
| 2. Psychological Reality = our knowledge and beliefs about 1. | B. Cognitive representations of this reality on the part of clinicians |
| 3. Propositions, Theories, Texts = formalizations of those ideas and beliefs | C. Publicly accessible concretizations of these cognitive representations in textual, graphical and digital artifacts |

45

Definitions



Entity = anything which exists, including things and processes, functions and qualities, beliefs and actions, documents and software (Levels 1, 2 and 3)

Domain = a portion of reality that forms the subject-matter of a single science or technology or mode of study;

Representation = an image, idea, map, picture, name or description ... of some entity or entities.

Representational Units = terms, icons, alphanumeric identifiers ... which refer, or are intended to refer, to entities; and do not have any proper parts which play this role

46

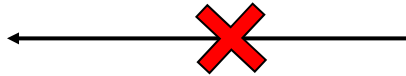
A representation is not the same as the entity it represents



CT Scan of the
Brain of Mr. X

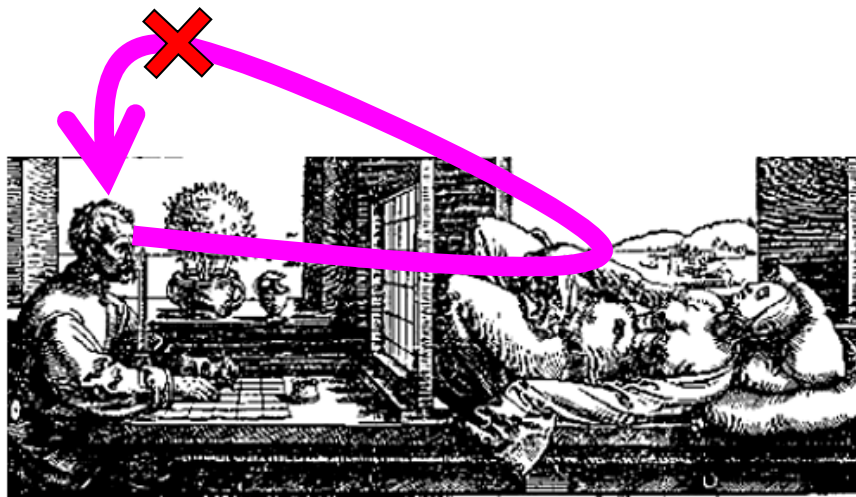
Brain of Mr. X

Ontology



47

Ontologies do not represent concepts in people's heads



48

So, an Ontology ...



- **Ontology** = a representational artifact whose representational units (drawn from a natural or formalized language) are intended to represent
 - types [of entities] in reality
 - those relations between these types which are true universally (= for all instances)

*lung **is_a** anatomical structure*
*lobe of lung **part_of** lung*

49

Results in ...



A tension between computer scientists and philosophers.

Philosopher's view: If the Ontology is built to represent *reality* then the exchange formats and data models based on it always remains valid allowing interoperability and ... and ...

Computer scientist's view: KISS

50

Results in the need to distinguish

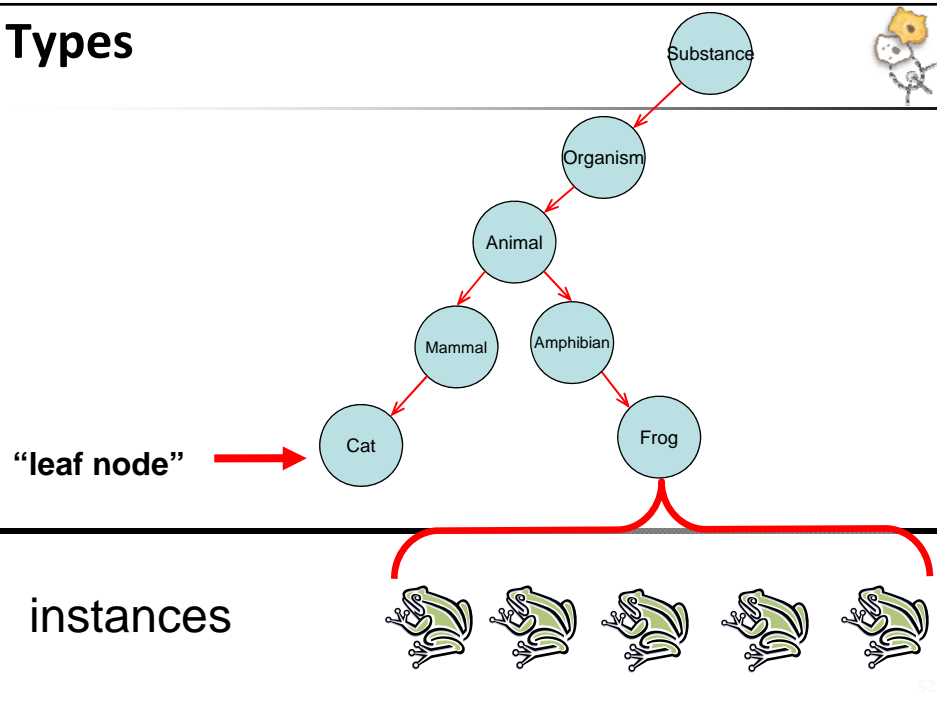


Ontologies, terminologies, catalogs: represent what is general in reality = **types** [classes]

Databases, inventories: represent what is particular in reality = **instances**

51

Types



Classes (Types) & Defined classes (Fiat types)



Class = a maximal collection of particulars determined by a general term ('cell', 'oophorectomy' 'VA Hospital', 'breast cancer patients in VA Hospital')

- **the class A** = the collection of all particulars x for which ' x is A ' is true

Defined Class = A class defined by a general term which does not designate a type in reality

- e.g. pathways

53

types < defined classes < 'concepts'



- Not all of those things which people like to call 'concepts' correspond to defined classes
- "Surgical or other procedure not carried out because of patient's decision" is a concept in SNOMED ...

54

Ontologies that represent concepts tend to make mistakes



1. congenital absent nipple **is_a** nipple concepts do not stand in
2. failure to introduce or to remove other tube or instrument **is_a** disease part_of
connectedness
causes
treats ...
3. bacteria **causes** experimental model of disease relations to each other

55

A Terminology is ...



A representational artifact whose representational units are natural language terms (with IDs, synonyms, comments, etc.) which are intended to represent defined classes.

Most Medical “Ontologies” are terminologies

56

The International Classification of Diseases



724 Unspecified disorders of the back
 724.0 Spinal stenosis, other than cervical
 724.00 Spinal stenosis, unspecified region
 724.01 Spinal stenosis, thoracic region
 724.02 Spinal stenosis, lumbar region
 724.09 Spinal stenosis, other
 724.1 Pain in thoracic spine
 724.2 Lumbago
 724.3 Sciatica
 724.4 Thoracic or lumbosacral neuritis
 724.5 Backache, unspecified
 724.6 Disorders of sacrum
 724.7 Disorders of coccyx
 724.70 Unspecified disorder of coccyx
 724.71 Hypermobility of coccyx
 724.71 Coccygodynia
 724.8 Other symptoms referable to back
 724.9 Other unspecified back disorders

57

ICD9 (1977): A Handful of Codes for Traffic Accidents



```

CiaM - icd9-cm.cla
File Edit View Class Modifier! Tools Options Help
New... Open... Print... Find... Centre
topClass
+D Diseases
+E Supplementary classification of external causes of inju
+E80 Railway accidents
+E81 Motor vehicle traffic accidents
+E82 Motor vehicle nontraffic accidents
+E820 NONTRAFFIC ACCIDENT INVOLVING MOTOR-DRIVEN SNOW V
+E821 NONTRAFFIC ACCIDENT INVOLVING OTHER OFF-ROAD MOTO
+E822 OTHER MOTOR VEHICLE NONTRAFFIC ACCIDENT INVOLVING
+E823 OTHER MOTOR VEHICLE NONTRAFFIC ACCIDENT INVOLVING
+E824 OTHER MOTOR VEHICLE NONTRAFFIC ACCIDENT WHILE BOA
+E825 OTHER MOTOR VEHICLE NONTRAFFIC ACCIDENT OF OTHER
+E826 PEDAL CYCLE ACCIDENT
+X E826.0 PEDAL CYCLE ACCIDENT INJURING PEDESTRIAN
+X E826.1 PEDAL CYCLE ACCIDENT INJURING PEDAL CYCLIST
+X E826.2 PEDAL CYCLE ACCIDENT INJURING RIDER OF ANIMAL
+X E826.3 PEDAL CYCLE ACCIDENT INJURING OCCUPANT OF ANIM
+X E826.4 PEDAL CYCLE ACCIDENT INJURING OCCUPANT OF STRE
+X E826.8 PEDAL CYCLE ACCIDENT INJURING OTHER SPECIFIED
+X E826.9 PEDAL CYCLE ACCIDENT INJURING UNSPECIFIED PERS
+E827 ANIMAL-DRAWN VEHICLE ACCIDENT
  
```

58

ICD10 (1999): 587 codes for such accidents



The image shows a large, dense table of ICD10 (1999) codes for accidents. The table is organized into columns and rows, with codes listed in a standard font. A red rectangular box highlights a specific section of the table, containing three entries:

- V31.22 Occupant of three-wheeled motor vehicle injured in collision with pedal cycle, person on outside of vehicle, nontraffic accident, while working for income
- W65.40 Drowning and submersion while in bath-tub, street and highway, while engaged in sports activity
- X35.44 Victim of volcanic eruption, street and highway, while resting, sleeping, eating or engaging in other vital activities

The table continues with many other codes, but they are not highlighted.

59



RELATIONSHIPS

The “is_a” relation



- What does *A is_a B* mean?
 - (A and B are types)
- For all x, if x **instance_of** A then x **instance_of** some B
- *cell division is_a biological process*

ALL-SOME STRUCTURE

61

The “part_of” (vs. has_part) relation



- | | |
|--------------------------------------|---|
| ▪ Human being has_part testis? | <i>A part_of B = all instances of A are instance-level parts of some instance of B</i> |
| ▪ human testis part_of human being ? | |
| ▪ Human being has_part heart? | <i>human testis part_of human being</i> |
| ▪ human heart part_of human being ? | |

62

Two kinds of parthood



between instances:

Mary's heart **part_of** Mary
this nucleus **part_of** this cell

between types

human heart part_of human
cell nucleus part_of cell

Copyright Stanford University 2006

63

The “part_of” relation



- What does A **part_of** B mean?
- For all x, if x **instance_of** A then there is some y, y **instance_of** B and x **part_of** y
 - where ‘**part_of**’ is the instance-level part relation
- *cell nucleus part_of cell*

ALL-SOME STRUCTURE

64

A part_of B, B part_of C ...



The **all-some structure** of the definitions allows cascading of inferences

1. within ontologies
2. between ontologies
3. between ontologies and EHR repositories of instance-data

65

Logical properties matter ...



- Expectations of symmetry may hold only at the instance level

- if *A* interacts with *B*, it does not follow that *B* interacts with *A*

- if *A* is expressed simultaneously with *B*, it does not follow that *B* is expressed simultaneously with *A*

Properties of Relations

1. Transitivity
2. Symmetry
3. Reflexivity
4. Anti-Symmetry
5. ...

67

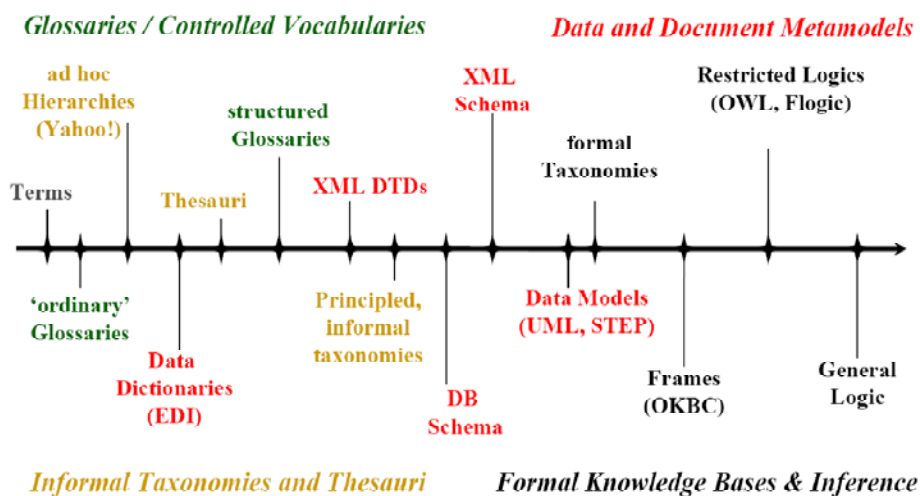
Other Ontology-like things



- **Controlled vocabulary** = A list of explicitly enumerated unambiguous terms; Controlled by a central registration authority;
- **Taxonomy** = collection of controlled vocabulary terms organized into a hierarchy
- **Thesaurus** = Collection of controlled vocabulary terms organized into a specialized network

68

Increasing “formality” ...



Originally by Michael Uschold, with permission

69

Application vs. Reference Ontologies



- A reference ontology is analogous to a scientific theory.
 - ... consists of representations of biological reality which are correct according to our current understanding.
- An application ontology is a software artifact:
 - ...for, structuring data according to some hierarchy of classes, for the purpose of managing, integrating and manipulating that data.
- As far as possible, we should focus on developing [scientific] information models, data-models, process-models etc to be as close as possible to and refer to *reference ontologies*.

70

Languages [formalisms] for Ontologies



- There are numerous ways of declaring both reference and application ontologies
- Almost all ontology languages give you the ability [and syntax] for declaring entities and relationships
- The main differences are in the ability [and mechanism] of describing the attributes of the entities and the mathematical properties of the relationships.
 - <http://xml.coverpages.org/OntologyExchange.html>
- Another major difference is the level of tool support available for “writing” in that language.
 - http://xml.com/2002/11/06/Ontology_Editor_Survey.html

71

Alternatives?



	Logical formalism	Reasoners	Tools to “speak” the language	Size of the user community	Status
OWL	Description Logic	Fact++ Pellet, Racer	Protégé, Swoop	~ in thousands	W3C standard
OBOF	DL-compatible for now	OBO-edit reasoner	OBO-edit	~ in hundreds	Bio* community standard
Frames (GFP) & OKBC	?	?	Ocelot, Ontolingua, GKB-Editor	~ in hundreds	AI community standard
KIF	expression of arbitrary logical sentences	?	?	?	AI community standard
Loom	Not DL	Loom “classifier”	Loom	Small	?
XOL					
SHOE					
OML					
RDF(S)	Subsumed by OWL	Subsumed by OWL	Subsumed by OWL	Subsumed by OWL	Subsumed by OWL
DAML + OIL	Subsumed by OWL	Subsumed by OWL	Subsumed by OWL	Subsumed by OWL	Subsumed by OWL

What an Ontology is NOT



- An ontology **is not** the same as **a knowledgebase**
 - Ontology (types) + Instances = KB
- An ontology **is not** the same as **a database schema**
 - A database schema is designed to store the instances conforming to an ontology
- An ontology **is not** the same as **an XSD**
 - An XSD tells you *how* to store the information that describes the instances

73